

机器如何学会打游戏

从塞缪尔的西洋棋开始，游戏就是机器学习里的研究热点，不仅是因为游戏的对战性可以直接体现出机器的智能，而且打游戏的技巧确实是可以学出来的，很适合用机器学习方法来解决。然而，在 2016 年以前，可能没有人会想到机器玩起游戏来竟如此强大，不仅可以在简单游戏中战胜人类业余选手，还可以在极为复杂的任务中战胜人类顶尖高手。

突破从 DeepMind 公司利用深度 Q-learning 网络（DQN）教会机器玩 Atari 游戏开始。Atari 平台包括多个小游戏，玩家通过控制杆来操作，完成特定任务后玩家会获得一定奖励，游戏结束后获得奖励最高者获胜。DeepMind 的做法是把游戏画面传给计算机，让它通过观察这些画面来控制游戏杆，像人一样操作游戏。问题是，如何训练这个机器人玩家，让它成为游戏高手呢？

最直接的思路是让机器向人学，记录人类玩家看到某一个画面时的操作方式，把这些记录作为训练数据来训练一个预测模型，依据这个模型，机器就知道看到某一画面应该采取什么样的动作了。这是一种典型的“监督学习”方法。这种方法看起来还可以，但要人做出各种场景下的示范还是很困难。更大的问题是，这种单纯向人类学习的方法无法突破人类技能的上限。

DeepMind 采用的是另一种学习方式：不告诉机器应该怎么做，只告诉机器这么做的收益，机器可随心所欲地发挥，只要最后收益最大就可以，这种学习方式称为“强化学习”。图 1 是强化学习的一个示意图。

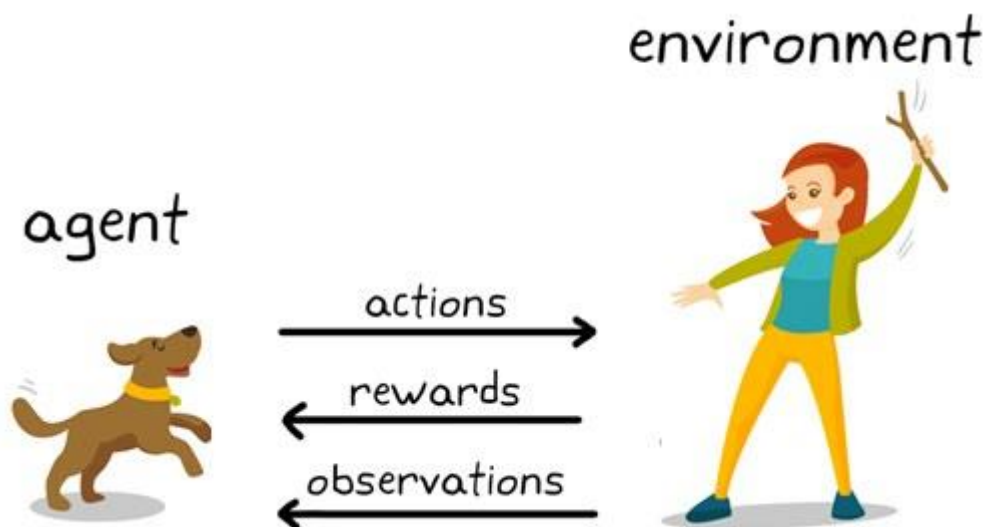


图 1：基于强化学习训练狗狗。狗狗做的好，主人给以奖赏，做的不好，给以惩罚。图片来自[2]。

强化学习在棋牌游戏中有广泛应用，但在 Atari 这种动作游戏上却比较困难，因为机器需要从这游戏画面中判断出游戏状态，这还是很困难的。为此，DeepMind 的研究员们用一个多层卷积神经网络（CNN）来提取状态信息，并利用这些信息直接预测机器应该采取的动作，

如图 2 所示。基于这一网络，机器每次选择潜在收益最大的动作来操作游戏杆，并将操作过程中得到的收益或惩罚作为神经网络的学习信号，利用 BP 算法对神经网络进行改进，直到成为游戏高手。

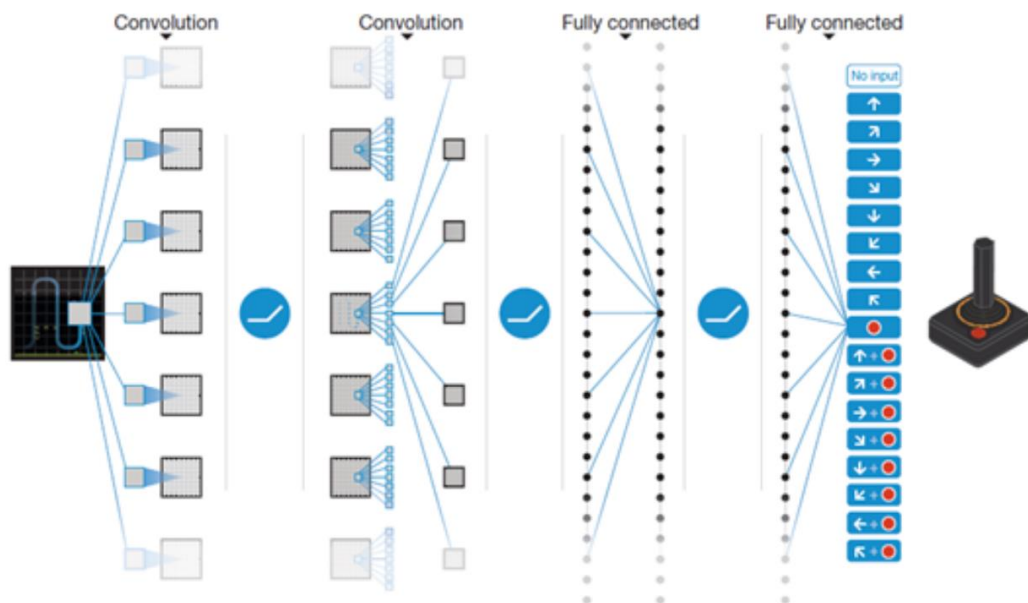


图 2：DeepMind 的 Atari 游戏控制程序，输入为游戏画面，输出为游戏杆的操作命令。

基于上述网络结构和强化学习方案，DeepMind 让机器自学了打 Atari 游戏的技能，在 50 款游戏的 29 款中取得了超过人类的性能，如图 3 所示。

深度强化学习在 Atari 游戏上的成功给了人们极大信心。后来，研究者提出各种新模型，在射击类、赛车类、策略类游戏上攻城略地，取得了一个个辉煌战果。图 4 是目前已经被机器尝试过的游戏及所采用的算法[3]。如果有一天你玩游戏时发现对手非常强大，也许对方就是机器也说不准。

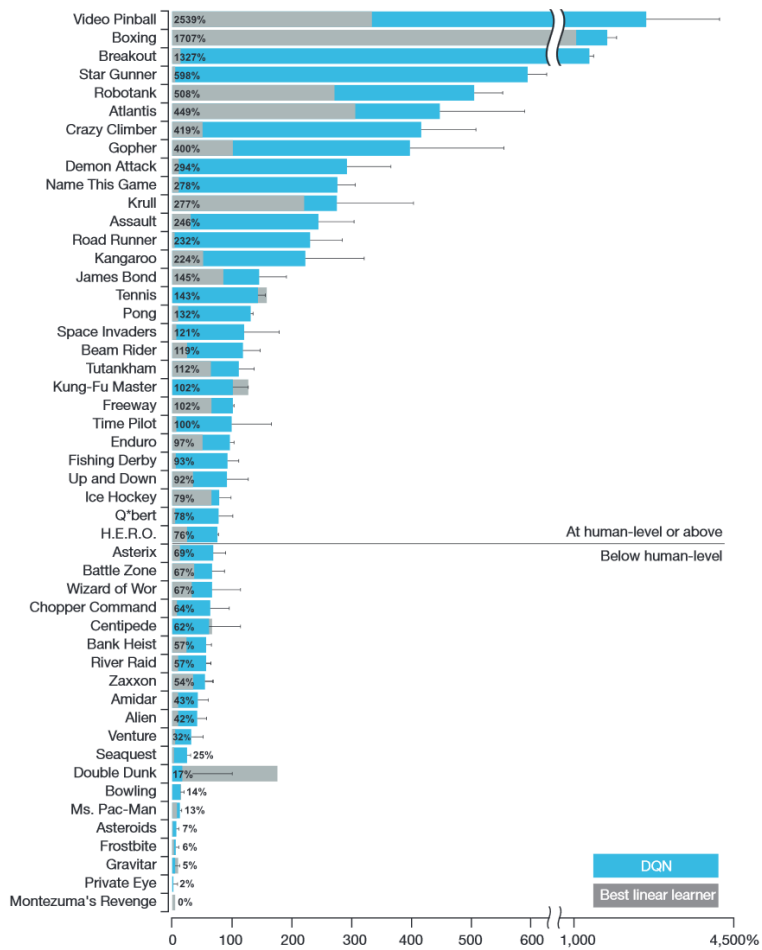


图 3: 基于 DQN 在 49 款游戏上的性能测试, 在 29 款游戏上超过了人类。

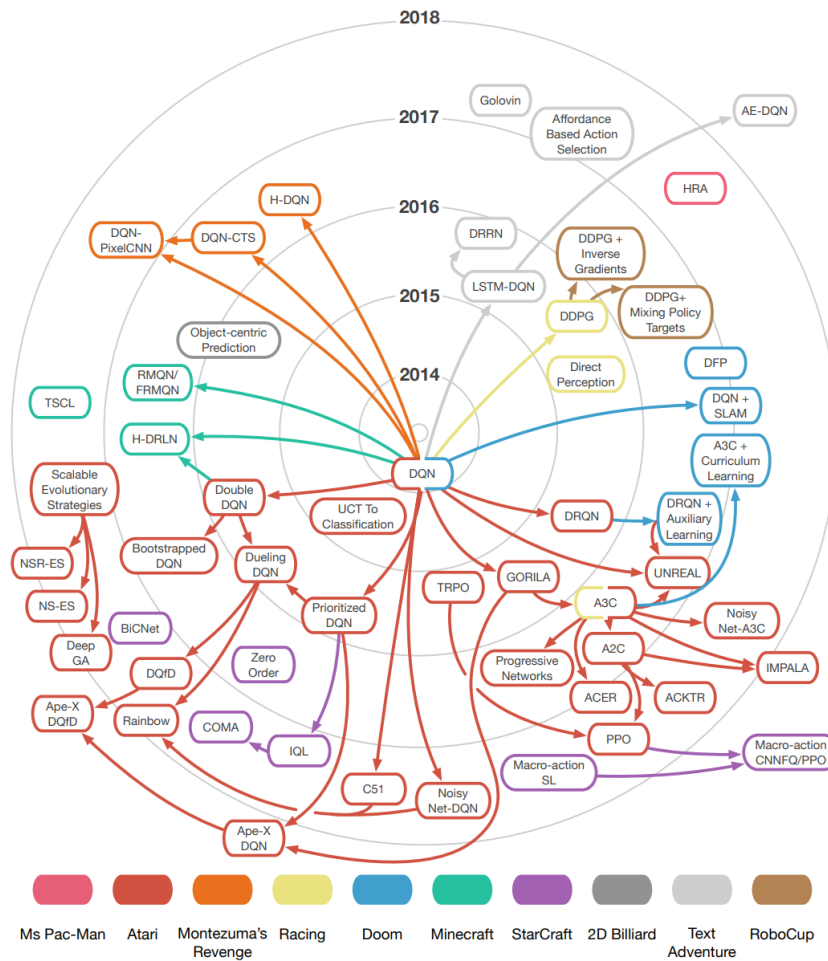


图 4: 被机器尝试过的游戏和相应的算法[3]

[1] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, et al. (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533

[2] Reinforcement Learning, <https://www.newworldai.com/reinforcement-learning/>

[3] Niels Justesen et al., Deep Learning for Video Game Playing, 2019.